



Learning of fuzzy spatial relations between handwritten patterns

Adrien Delaye, Eric Anquetil

► To cite this version:

Adrien Delaye, Eric Anquetil. Learning of fuzzy spatial relations between handwritten patterns. International Journal of Data Mining, Modelling and Management, 2014, 6 (2). hal-00735545

HAL Id: hal-00735545

<https://inria.hal.science/hal-00735545>

Submitted on 26 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning of fuzzy spatial relations between handwritten patterns

Adrien Delaye* and Eric Anquetil

INSA de Rennes - UMR IRISA
 Campus de Beaulieu, F-35042 Rennes,
 Université Européenne de Bretagne, France
 adrien.delaye@irisa.fr - eric.anquetil@irisa.fr
 *Corresponding author

Abstract:

It is widely admitted that modeling of spatial information is very important for interpretation and recognition of handwritten expressions. Two distinct tasks have to be addressed by spatial models in this context. *Evaluation task* consists in measuring the correspondence between the relationship of two objects and a predefined model of spatial relation. *Localization task* consists in retrieving objects that are related to a reference object according to a predefined model of spatial relation. In this work, we introduce a new modeling of relative spatial positioning that handles the two tasks under a unified framework and a training scheme for learning spatial models from data. The use of fuzzy mathematical morphology allows to deal with imprecision of positioning and to adapt to varying shapes of handwritten objects. Experimentations of the evaluation task over two datasets of online handwritten patterns prove that the proposed modeling outperforms commonly used relative positioning features.

1 Introduction

In many computer vision related tasks, spatial relationships between entities convey an important part of the information and should be modeled and processed with special attention. Interpretation of complex hand-drawn schemes involves the analysis of structured two-dimensional patterns where spatial relations between components provide useful information about the global meaning. For example, when dealing with handwritten mathematical equations, one has to analyze the relative positioning of isolated symbols in order to interpret the global meaning of the equation. In this specific case, a semantic (mathematical) operator such as *exponent* is only expressed by the *superscript* spatial relation between its two operands.

Structural representations are commonly used for description and analysis of complex expressions. Whether they rely on graphs, trees, grammars or other concepts, they involve description of spatial relations or connections between structural components. In this work, we are interested in the description and modeling of pairwise relations between components of handwritten structures,

but do not consider integration of spatial relation descriptions into higher-level structural representations.

While human beings are able to intuitively think of relative positioning with the help of concepts such as distance (*far from, close to...*) or directions (*on the left of, above...*), it is clear that these concepts are not precise by nature, and that they require a flexible modeling supporting reasoning under imprecision and dealing with ambiguity. Actually, applying soft computing concepts for modeling spatial relations is a widely accepted idea, and many methods have been proposed in this direction (see for example the books [MS02, JPPS10]). In the context of hand-drawn patterns processing, it seems natural to make use of these concepts for handling objects that are by nature noisy, imprecise, and subject to a strong variability according to the numerous input conditions (writer identity, nature of the input material, environment, space and time...). Whereas much research effort has been spent on modeling and recognition of isolated shapes [PS00], most of spatial positioning models rely on ad hoc measures that do not precisely take into account variations in the shapes of objects and thus lead to poorly intuitive and informative descriptions.

In this work, we introduce a new formalization of spatial relation models based on fuzzy mathematical morphology operators. The original idea, formulated by Bloch in the context of MRI brain imaging [Blo99], provides a description of relative positioning that perfectly adapts to the variations in object shapes. In previous work, we highlighted the benefits of considering shape variations when describing relative positioning, in the context of handwritten recognition [BMA06]. The main contribution of this work is to introduce trainable models of spatial relations that can address the two important tasks related to spatial analysis of structured expressions, namely the *evaluation task* and the *localization task* [MWN10]:

- Evaluation of a spatial relationship between two objects A and B consists in qualifying how well their relationship matches the spatial relation M (where M is a model of spatial relation either predefined from expert knowledge or estimated from data). For analysis of complex expressions, it gives an evaluation of spatial relations hypotheses and thus participates in the global interpretation.
- Localization consists in finding objects that best satisfy a predefined or trained spatial relation M with respect to a given reference object R . For expression analysis, it guides the structural parsing by retrieving objects conforming to the spatial relation model with respect to R .

To the best of our knowledge, no existing method combines a direct handling of these two tasks with an automatic training process to learn models from data. Relying on the framework of fuzzy mathematical morphology, we present here a unifying formalization for handling the two aforementioned tasks, and a training scheme for building models from data. Localization task is performed by application of the models on a reference object which results in a description directly defined in the image space. The evaluation of a spatial relationship between two objects can be performed by application of the same model. In addition to proposition of new modeling of spatial relations, we also focus in this work on the integration of distance information in this framework.

Second section reviews methods for relative positioning of objects, either in the context of handwriting recognition or in the more general domain of image processing. Principles of using fuzzy morphological operators for describing spatial relations are then detailed in section 3. Formalization of new models and associated learning process is presented in section 4, and integration of the distance into the positioning models is discussed in section 5. Finally, experimental results obtained in classification of handwritten objects are presented in section 6, followed by concluding remarks and directions for future works.

2 Related work

This section reviews methods for modeling spatial relations between objects. Note that general representations of complex structures are not investigated here, since we limit the scope of this review to models of pairwise spatial relations between structural components. Methods for spatial relation description can be used in different structural representations of complex handwritten expressions or, for instance, in segmentation graphs for image understanding. We first briefly review methods dedicated to modeling spatial relations between handwritten objects. Then we investigate methods proposed for modeling of spatial relations in the domain of image analysis.

We mainly restrict our review to methods that adopt a fuzzy modeling of relative positioning. Since the use of soft representations for spatial reasoning was suggested by Freeman more than thirty years ago [Fre75], it has been widely applied and development of soft computing methods for spatial reasoning has still been a very active lead of research during the past decade [Blo99, MW99, Blo05, MWN10, MS02, JPPS10].

2.1 In the handwriting processing domain

Handwritten objects have several characteristics that make them worth distinguishing from general image objects. In general, handwritten objects are represented as *strokes* which are *linear* structures (i.e. that have no thickness), in opposition to *areal* objects dealt with in image processing domain. While much effort has been spent on precisely modeling the shapes of handwritten strokes, most models for describing their relative positioning rely on rather simple measures provided by ad hoc methods. Common approaches consist in summarizing each hand-drawn primitive by a single synthetic point (e.g. centroid based methods) or a virtual rectangle (minimum bounding rectangle or bounding box based methods) [MA04, AMVG10], thus sacrificing the actual shape of objects. If it is admitted that when two objects are sufficiently far from each other their positioning can be summarized by the relative positioning of their centroids, this simplification practically never holds in the context of strokes within a handwritten expression. Likewise, the bounding box approximation is too coarse for objects that have complex shapes, including singularities, concavities, loops, all of which are common characteristics of handwritten patterns.

To deal with evaluation task, most methods proceed by extracting positioning features from a pair of objects and comparing their adequacy to spatial relation

models defined in the feature space (models can be trained from data, or predefined from expert knowledge). This scheme is commonly used and strongly benefits from statistical classification tools that provide very efficient automated model learning. However, models defined in a feature space distinct from image space cannot be directly used for spatial localization tasks, because features are extracted from pairs of objects (the two objects have to be available for feature extraction, whereas for localization task only the reference object is available).

For handling localization task, existing methods make use of regions empirically defined in the image space. For example, *Zhang et al* define around a symbol several rectangular regions corresponding to available spatial relations in mathematical expressions (such as *superscript*, *subscript*, *below*...). Regions definitions are based on the bounding box of the reference symbol and their boundaries are parallel to the image axes. This approach is commonly used for mathematical expression analysis [ZBC02][KK10]. Fuzzy extension of region definitions can help handling imprecision of positioning, but it does not improve the adaptability to varying object shapes which remains minimal because of the bounding box approximation [ZBZ05]. More importantly, several thresholds have to be manually set for constructing the models and no scheme is proposed for automatically building models from samples of the spatial relations, hence a lack of genericity.

In general, positioning methods for handwritten objects are dedicated to very specific domains (e.g. mathematical expressions, Latin characters, Asian characters, ...). The bounding box approximation is widely used, and no generic method was proposed that takes into account the actual shapes of handwritten strokes for defining more precisely their relative positioning. On one hand, methods for evaluating spatial relations properly handle variability by training statistical models in a feature space, but are not suited to locate objects. On the other hand, solutions proposed for localization are empirical, require manual definition of spatial regions, and have trouble capturing the variability of spatial positioning.

2.2 In the image processing domain

In the domain of image analysis, much more researches were conducted with the aim of defining generic methods for modeling relative positioning of objects. A review of fuzzy-based methods for defining spatial relations in images was proposed by Bloch [Blo05]. Among these methods the ones that take into consideration the shapes of the objects involved in the relation, rather than operating a centroid based or box based approximations, are of much interest.

Several approaches provide a rich description of relative positioning by implicitly taking into account the shapes of the objects: they consider all the points from the two objects at hand for describing their spatial relation. This is the case for methods based on histogram of angles (compatibility method), or for the aggregation method [Blo05]. The F-histogram method also takes into account object shapes by aggregating positioning relations over longitudinal sections of the objects, and is computationally efficient in comparison with direct computation of histogram of angles [MW99]. The quality of descriptors extracted by histogram of angles or F-histogram was validated in several applicative contexts. Besides evaluation of spatial relations between objects, F-histograms have also been used

for generating linguistic description of relations [BMK04], or for retrieving similar spatial relationships among images [Mat02]. However, directly exploiting them for localization task is impossible, because their definition is based on features extracted from a pair of objects.

The authors of F-histogram more recently introduced another representation of spatial relations, dual to F-histogram, called F-template [WNM06]. An F-template is a spatial template defined relatively to a reference object, describing for each point of the plane its adequacy to a given spatial relation with respect to the reference. For instance, an F-template can describe the relation *to be in direction α from R* , where R is a reference object. A F-template is a type of model that can be used for localization of objects, since it is directly defined in the image space [MWN10] and only requires a reference object. The duality of F-histograms and F-templates concepts has not been fully investigated yet, and no direct transformation is possible from one to the other, but these works constitute an interesting step toward unified models for evaluation and localization tasks.

A similar concept of spatial template has been investigated for about ten years in the works of Bloch. The original idea, formulated separately by Bloch [Blo99] and Gader [Gad97] is to describe spatial relations with the help of morphological operations processed in the image space, directly over the reference object. Specifically, fuzzy functions that model predefined spatial relations are applied with the help of morphological operators (mainly dilatation) to build a spatial template (called *fuzzy landscape* in their terminology). Obtained spatial template is a fuzzy function describing the goodness of a spatial relation with respect to the reference, for any point of the space.

F-templates and morphological *fuzzy templates* both take into account the shape of the reference object in the description of the spatial relations. Objects shapes are even explicitly considered in the morphological method, and Bloch has shown that the resulting description better fit the intuition in comparison to other positioning descriptions based on histograms [BR03]. F-templates and morphological approach provide two different ways to build spatial templates, and Matsakis pointed out the equivalence of the two definitions under some conditions [MWN10]. Spatial templates effectively handle the localization task, since they are defined directly in the image space and only require the definition of one reference object. However, in the two cases, only *predefined* spatial relations can be represented (such as *on the left*, *above*, *close to*), and no method was proposed to automatically train such models from data.

In the context of handwriting expression processing, spatial relations are not easy to describe or to model from prior expert knowledge. The need of generic spatial relation models with respect to different types of expressions and objects calls for an automatic training procedure. Moreover, automated learning of spatial relations would be beneficial for better capturing the variability and imprecision of relative positioning of objects.

The main contribution of our work is to propose a new formalization and an associated procedure for learning models of spatial relations from training data. The result of training is an abstraction of the spatial relation (for example *superscript*). When applied to a given reference object, the model results in a spatial template of the relation in the image space that can be used for localization task. When a pair of objects is available, the learned model can be used for

evaluation of their positioning. The proposed approach is built upon the fuzzy morphological method, because it is flexible and perfectly suited to deal with ambiguous spatial relations. Morphological foundation is also precious for building spatial relation models that can handle a large variability in the shapes of the objects and perfectly adapt to their individual specificities. In the next section, we present the idea of using fuzzy mathematical morphology for describing of relative positioning and then introduce the new proposed modeling in section 4.

3 Relative positioning description with fuzzy mathematical morphology

We first present the general approach as introduced in [Blo99] for describing relative positions of objects. Although the method was designed to deal with fuzzy objects in images and can be extended to the 3D case, we only present here the concepts required for our target application: space is reduced to the plane, and we only consider crisp linear objects.

A spatial relation is modeled as a fuzzy set describing the adequacy of any point of the plane to the relation at hand relatively to a reference object. Such type of relation can be based on directional information (for example *on the right of R* , where R is the reference object) or distance information (*far from R*). The use of fuzzy models is natural, since these concepts cannot be described in a crisp, all-or-nothing fashion.

In the sequel, a fuzzy set describing a spatial relation in the plane is referred to as a *fuzzy landscape*, according to the vocabulary defined in [Blo99]. This concept matches the notion of *spatial template* referred to by Matsakis et. al [MWN10]. Intuitively, a fuzzy landscape can be seen as a *map* (think of a distance map, or a direction map) which associates to every point of the plane a degree of truth with respect to the relation described, and the chosen reference object. An example of fuzzy landscape describing the spatial relation *on the right* of an object R is presented in figure 1(b). In this figure, fuzzy membership values from 0 to 1 are represented by levels of gray from black to white (this convention will be kept for representing fuzzy landscapes throughout this paper). Object R is a primitive extracted from a handwritten Chinese character. Since Chinese characters are complex structured patterns, they present a rich set of examples of spatial relations. We choose to illustrate all the concepts and functions presented in this paper with objects taken from handwritten Chinese characters.

Evaluation of a positioning relation for a given object A with respect to the reference R involves two steps: first step consists in defining the *fuzzy landscape* by operating a morphological operation over the reference object R , and second step consists in evaluating the adequacy of object A with the *fuzzy landscape*.

3.1 Fuzzy landscape definition

First step consists in applying a predefined model of spatial relation, for example *on the right of*, to the reference object of interest, say R , in order to model the applied relation *on the right of R* . This is done by applying a morphological dilatation of the reference object with a structuring element (SE) modeling the

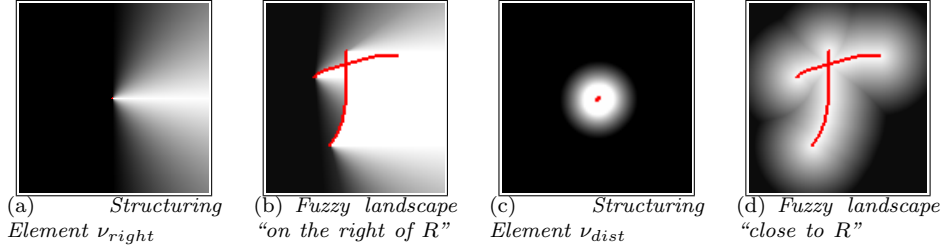


Figure 1 Structuring Elements for spatial relation *on the right of* (a), and *close to* (c), and resulting fuzzy landscapes around a reference object R (b, d). Brightness reflects membership to the fuzzy functions, from 0 (black) to 1 (white).

relation. For a given reference object R , located in the plane S , the dilatation of R with the SE ν is computed by:

$$\mu^R(p) = \max_{q \in R} \nu(p - q), \forall p \in S \quad (1)$$

(where p and q are points of the plane S , with q belonging to the object R).

The resulting function μ^R is the *fuzzy landscape*, and it describes for each point of the plane its adequacy degree to the relation described with respect to R , illustrating the applied relation. Its definition explicitly takes into account the shape of the reference object R . Note that compared to application in the image domain, the computation of (1) is much facilitated when dealing with linear objects, since R has significantly less points. Dealing with on-line signal (i.e. when the handwritten strokes are described as a temporal sequences of sampled points) is even faster, as the handling of linear sections is straightforward and computationally efficient, as shown in [BMA06]. In this case, the computational effort for evaluating μ^R at a point p is linear with the number of sampled points in the reference object.

The principle of equation (1) is to look for the point of R that best supports the relation held by ν for the point p . Consequently, a point will be considered to be perfectly *on the right of* R if p is perfectly on the right of any part of R . As a result, one point p can be at the same time perfectly on the right of (some part of) R and perfectly on the left of (some other part of) R . What can be seen as a loss of information (in comparison to methods like histograms of angles where all pairs of points from the two objects are used for describing their relationship) actually permits to get a precise knowledge about the shape of R by considering several directions. For example, a point p located inside the concavity of a C-shaped reference object is simultaneously perfectly *above*, *under* and *on the right of* R . This idea of combining several points of view for building rich descriptions of relative positioning is at the core of the models introduced in the following section.

3.2 Structuring element definition

The shape and the fuzziness of SE ν define the spatial relation and the softness of the description provided by the fuzzy landscape (resulting from equation (1)). As

an example, the following definition can be adopted for the fuzzy SE modeling the spatial relation *in direction* α [Blo99]:

$$\nu_\alpha(p) = \max \left(0, 1 - \frac{2}{\pi} \arccos \frac{\vec{op} \cdot \vec{u}_\alpha}{\|\vec{op}\|} \right), \forall p \in S \quad (2)$$

where o is the center of the SE and u_α is the unit vector of direction α . SE fuzzy membership function ν assigns to each point p of the plane a degree from 0 to 1, varying linearly with the angle between \vec{op} and u_α . Figure 1(a) shows a SE for the relation *on the right of* defined according to equation (2) with $\alpha = 0$, and figure 1(b) illustrates the fuzzy landscape obtained by dilatation of a reference object with this SE, according to equation (1).

Alternatively, an undirected fuzzy radial SE can be defined for modeling a distance relation (e.g. *close to*). Figures 1(c) and 1(d) present an example of distance based SE and resulting fuzzy landscape for an object R . Different definitions of structuring elements were proposed in [CA07], where the authors introduce several parameters controlling the flexibility of the radial and directional dimensions to provide a customized family of hybrid distance-direction fuzzy structuring elements.

3.3 Evaluation of the relationship

Once the fuzzy landscape is defined for a spatial relation (either based on directional, distance-based, or hybrid structuring elements), it can be used for evaluating the positioning of an object A . Several operators are available for aggregating membership degrees of points of A to the fuzzy landscape so as to obtain a global evaluation. Mean measure simply computes an average adequacy degree (3), while possibility (4) and necessity (5) respectively provide optimistic and pessimistic measures that define an evaluation interval according to the fuzzy pattern matching theory [Blo99]:

$$M^R(A) = \frac{1}{|A|} \sum_{p \in A} \mu^R(p) \quad (3)$$

$$\Pi^R(A) = \sup_{p \in A} \mu^R(p) \quad (4)$$

$$N^R(A) = \inf_{p \in A} \mu^R(p) \quad (5)$$

As an illustration, table 1 gives the three degrees measured for the object A embedded in the 4 directional fuzzy landscapes depicted in figure 2. For example, according to direction $\alpha = 0$, the mean score of 0.87 shows that object A is mostly on the right of R . The two other measures reflect the fact that at least one point of A is *perfectly* on the right of R in the sense of the fuzzy landscape (hence possibility is equal to one), while at least one point of A is not really on the right of R (hence necessity is 0.6). Because they only rely on one point, measures of necessity and possibility are quite sensitive to outliers, but they provide complementary information about the distribution of degrees reached by points of A .

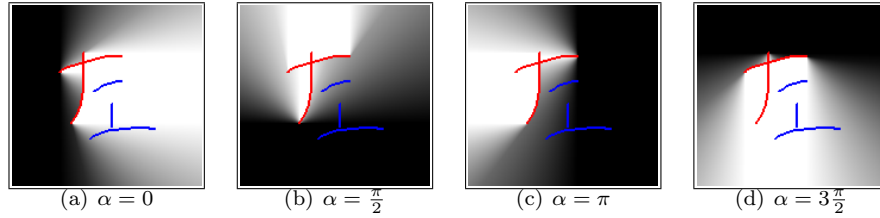


Figure 2 *Is A in direction α from R?* (reference object R is in red, argument object A is in blue)

Table 1 Evaluation of the relations *Is A in direction α from R?* (for objects of figure 2)

| Measure | $\alpha = 0$ | $\alpha = \frac{\pi}{2}$ | $\alpha = \pi$ | $\alpha = 3\frac{\pi}{2}$ |
|------------|--------------|--------------------------|----------------|---------------------------|
| $M^R(A)$ | 0.87 | 0.05 | 0.03 | 0.93 |
| $\Pi^R(A)$ | 1 | 0.2 | 0.15 | 1 |
| $N^R(A)$ | 0.6 | 0 | 0 | 0.77 |

4 Learning unifying positioning models

As exposed in the previous section, use of mathematical morphology permits to take into account the actual reference object when defining the fuzzy landscape, so that spatial relation description naturally adapts to the peculiarities of its shape. The fuzzy landscape is suitable to a human-like description of spatial relations and fits the intuition very well, including when the reference object has singularities, concavities... Moreover, thanks to its fuzzy definition, it deals well with the imprecise relative positioning of handwritten patterns and the spatial relationship evaluation can be integrated in a general soft computing framework for making recognition decision. Finally, since a fuzzy landscape is a spatial template (i.e. a model of spatial relation that can be represented in the image space), it can be used for addressing the localization task. The several evaluation measures presented offer different strategies for evaluating the matching of a spatial relationship between two objects with respect to a predefined model.

Nevertheless, this method only describes predefined linguistic spatial relations (e.g. *on the right of*, *close to*). In previous works, we have used combination of several linguistic fuzzy landscapes to evaluate how well the relationship between objects fits with different directional spatial relations. Evaluation results were shown to constitute a powerful set of features for classification of spatial relations between handwritten strokes [BMA06].

Our objective is to further exploit the fuzzy morphological framework by automatically learning abstract modeling of spatial relations that will provide learned fuzzy landscapes when applied to a reference object. This modeling allows to cope with both localization and evaluation tasks:

1. given a reference object and this spatial model, in what area of the plane should the argument object be located? (localization by definition of the learned fuzzy landscape)

2. given two objects, to what extent does their spatial relationship fit with the model? (evaluation by matching of the argument object with the learned fuzzy landscape)

Learning of the spatial model should not require any prior knowledge on the spatial relation, such as preferred directions or distance. It should also handle large variety of spatial relations, including different topological situations, with intersecting objects as well as surrounding and inclusion situations.

In this section, we present a method for automatically learning from samples spatial relation models by relying on fuzzy morphology operators. We propose to examine the spatial relation under 4 different *points of view* (the 4 cardinal directions). For each of them, a partial model can be determined, applicable as a modified fuzzy landscape for describing the area admitted by the model considering this point of view. Global model is ultimately constructed by fuzzy intersection of modified fuzzy landscapes obtained according to the several points of view. Introduction of distance information in the model will be investigated in a dedicated section (section 5).

4.1 Partial directional model

From several training instances of objects $(R_i, A_i)_{i=1..N}$ that share the same spatial relation to be modeled, we learn a partial model of their positioning according to each of the 4 adopted points of view, oriented by *up*, *down*, *left* and *right* directions. For a point of view related to direction α , we consider the distribution of the degrees reached by the training points from objects A_i with respect to the directional fuzzy landscape built relatively to their associated reference objects R_i . We simply approximate this distribution by a histogram function H_α , normalized by the maximum frequency. For $k = 0..K$ and $(i_k)_{k=0..K}$ in $[0, 1]$ such that $i_k = k/K$, each histogram section $(H_\alpha)_k$ accounts for the number of points from the training samples that reach a degree falling in $[i_k, i_{k+1}[$ in the fuzzy landscape $\mu_\alpha^{R_i}$:

$$(H_\alpha)_k = \sum_{i=1..N} |S_{i_k}| \quad \text{where} \quad S_{i_k} = \{p \in A_i, \mu_\alpha^{R_i}(p) \in [i_k, i_{k+1}[\} \quad (6)$$

After a simple normalization so as to set the maximal histogram value to 1, (H_α) is a function from $[0, 1]$ to $[0, 1]$ approximating the distribution of degrees reached by training points with respect to the relation carried by α . H can be interpreted as a model of *to what extent* A_i objects are in direction α with respect to R_i objects in the training set. Figure 3(a) presents a histogram modeling the distribution of degrees reached by points of objects similar to the (R, A) pair of figure 2, considering the direction *on the right of*.

Exploitation of model H_α consists in functional composition of the linguistic directional fuzzy landscape with the function H_α :

$$\hat{\mu}_\alpha^R = (H_\alpha \circ \mu_\alpha^R) \quad (7)$$

The result of composition is a modified fuzzy landscape $\hat{\mu}_\alpha^R$, interpretable as a spatial model of the relation *to be located with respect to R according to the model, under the point of view α* .

For example, figure 3(b) shows a learned fuzzy landscape representing the spatial relation between objects R and A from figure 2 according to the point of view *right*. It assigns to each point of the plane its adequacy degree with the learned relation according to this direction and can be interpreted as *to be conformely to the model "on the right of R "*. In other words, the brighter points are those for which the validity of relation *to be on the right of R* is the most conform to the degrees reached by the training objects. The white area could be used to localize objects, i.e. to predict where it is expected to find an object positioned relatively to this given reference object, according to the model learned and the considered point of view (*right*).

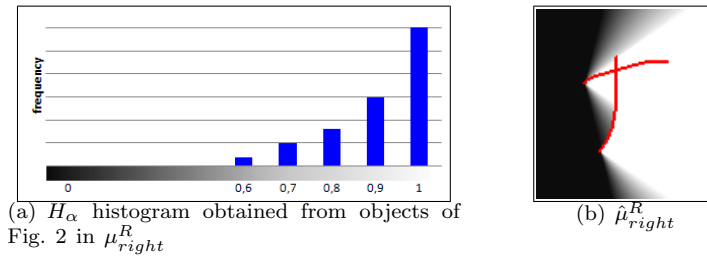


Figure 3 Histogram model (a) and resulting learned fuzzy landscape (b) for point of view *right*, with training objects similar to the (R, A) pair from figure 2

4.2 Fusion by fuzzy intersection

The process described in the previous part is repeated for each considered point of view, i.e. for the 4 directions *up*, *down*, *left*, *right*. When given a new reference object, the four models are exploited separately, resulting in four distinct learned fuzzy landscapes forming sub-parts of a global spatial model. This global model is then constructed by intersecting the fuzzy landscapes obtained by considering each point of view:

$$\hat{\mu}_{inter}^R = t[\hat{\mu}_{up}^R, \hat{\mu}_{down}^R, \hat{\mu}_{left}^R, \hat{\mu}_{right}^R] \quad (8)$$

where t is a t-norm fuzzy operator, implementing a conjunction operation. Intuitively, a point p is considered as properly positioned with respect to the model if it fits with the learned landscape according to every point of view.

Figure 4 illustrates the fuzzy landscapes obtained for the 4 points of view (a, b, c, d), with the same objects as before. In each case, the white area describes the admitted region according to the model under the considered point of view. The two fuzzy landscapes obtained for points of view *up* (a) and *down* (b) are partly redundant. Information extracted from the *up* point of view is that A should be not above R , while information from the *down* point of view is that A should be below R . These two points of view actually complement each other, and the fuzzy landscape from the *up* point of view is not completely included (in the fuzzy inclusion sense) in the one from *down* point of view. It is then beneficial

to intersect these two landscapes for building a more accurate global model. The same reasoning holds for *left* and *right* points of view (c, d). Eventually, the global model computed by intersection is represented in figure 4(e). This global landscape $\hat{\mu}_{inter}^R$ describes the area of the plane that satisfies the positioning relation with respect to R according to the complete model.

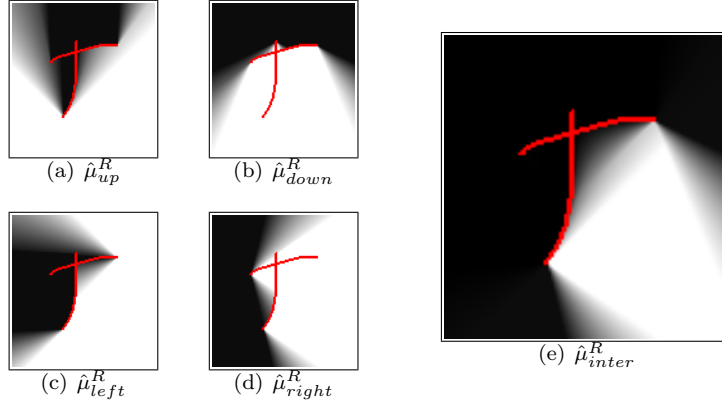


Figure 4 Exploitation of a spatial relation model trained from objects such as the (R, A) pair of figure 2. Partial learned fuzzy landscapes are obtained from the 4 points of view (a), *down* (b), *left* (c) and *right* (d). Each point of view provides a different information regarding the admitted location of an object with respect to the reference R . The global landscape (e), obtained by intersection of partial models, is a spatial representation of adequacy to the learned spatial relation model.

A global model defined by $(H_{up}, H_{down}, H_{left}, H_{right})$ is an abstraction of a non-linguistic spatial relation. Given a new reference object R , application of the model results in a fuzzy landscape that is adapted to the shape of the reference. As in the case of predefined spatial relations, representing learned relations with fuzzy landscapes unifies the needs for localization and evaluation task. This learned fuzzy landscape can directly be used for localization task, since it exhibits the area of the plane where argument objects should be looked for. A second exploitation level assesses the matching of the relationship between A and R with respect to modeled spatial relation. Evaluation of the matching of A with the fuzzy landscape defined on R can be seen as an evaluation of the proposition *the relative positioning of R and A matches the learned spatial relation model*.

Figures 5(a) to 5(d) present different fuzzy landscapes obtained by application of the same spatial relation model to several reference objects. Figures 5(e) to 5(h) show the matching of associated argument objects to the fuzzy landscapes and corresponding mean evaluation measures.

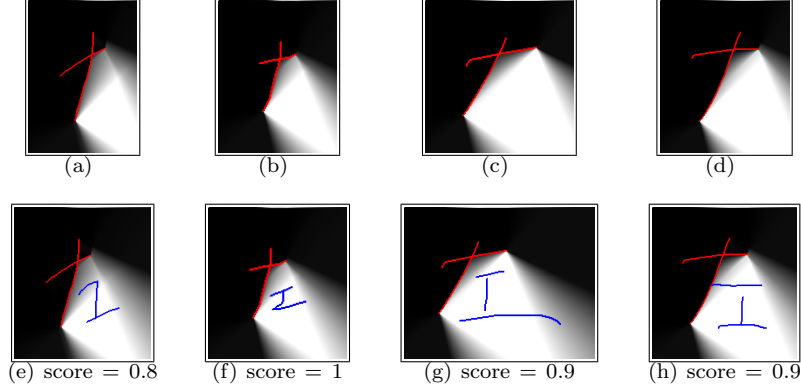


Figure 5 Application of a learned spatial relation model to different pairs of objects. Fuzzy landscapes adapted to the reference objects (a)-(d) and associated argument objects (e)-(h) with corresponding mean adequacy scores.

5 Integration of distance information

The models presented in the previous section only involve directional positioning information. The combination of several points of view (i.e. consideration of several directional relations) provides a rich positioning description because different parts of the reference object support the different directional relations. It is however natural to investigate the contribution of distance information for improving the description quality. In this section, we consider two methods for embedding information about the distance between objects in the models. First method is the most straightforward: it simply consists in considering the distance as an additional point of view on the spatial relation. Alternative method associates a distinct distance model to each considered direction. The two methods will be compared in the experimental section (see section 6).

5.1 Global distance integration

The most straightforward way to model distance information within our framework is to make use of distance-based fuzzy structuring elements, and to consider the distance as an additional point of view in our model. We propose to define a fuzzy distance-based structuring element as follows:

$$\nu_\tau(p) = \beta_\tau(\|\vec{op}\|) \quad \text{with } \beta_\tau(t) = \max\left(0, 1 - \frac{t}{\tau}\right) \quad (9)$$

(for any point p of the plane). o denotes the center of the structuring element, and β is a decreasing function from \mathbb{R} to $[0, 1]$, that has a null value when $t > \tau$. It actually defines a model of the spatial relation *close to*. A distance fuzzy landscape μ_{dist} is obtained by dilatation of a reference object R by the distance structuring element. In order to adapt the distance relation to the size of the reference object,

the scaling factor τ is set proportionally to the length of the diagonal of the bounding box of R . The distance fuzzy landscape is defined as

$$\mu_{dist}^R(p) = \max_{q \in R} \nu_{\tau R}(p - q), \forall p \in S \quad (10)$$

The model can be defined and trained just as explained before. Figure 6(a) shows the learned fuzzy landscape $\hat{\mu}_{dist}^R$ obtained for distance point of view. It gives a global representation of acceptable regions of the plane in terms of distance to the reference object, according to the training data. Distance is modeled globally, without considering differences with respect to the directions from R . Exploitation of the global model now involves the additional distance fuzzy landscape in the fuzzy conjunction for combining points of view (see equation 8). Result of the intersection of the 5 points of view (4 directions + 1 distance) is given in figure 6(b), and it can be compared with figure 4(e) (where no distance was included).

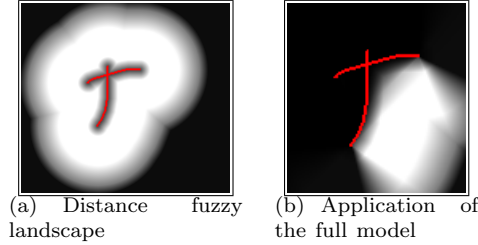


Figure 6 A fuzzy landscape for distance modeling can be seen as an additional point of view on the relation (a). The result of the global model, obtained by intersection the 5 points of view is presented in (b).

5.2 Direction-wise distance integration

Another method for embedding distance information in the models is to consider different distance models depending on the directions. It is expected that a more accurate modeling of the distance can be reached by incorporating a distinct model into each considered point of view. The idea is then to learn a distance-based fuzzy set for each histogram bar in the models introduced in section 4. Formally, if $(H_\alpha)_k$ denotes the k -th bar of the histogram model for directional point of view α , associated distance model $(\delta_\alpha)_k$ should approximate the distribution of distance degrees reached by point of S_{i_k}

$$(\Delta_\alpha)_k = \{\mu_{dist}^R(p), p \in S_{i_k}, i = 1..N\} \quad S_{i_k} = \{p \in S, \mu_\alpha^R(p) \in [i_k, i_{k+1}]\}$$

where i indexes a pair (R_i, A_i) of training samples, and μ_{dist}^R is a distance fuzzy landscape defined as exposed in the previous paragraph. $(\Delta_\alpha)_k$ is approximated by a fuzzy set $(\delta_\alpha)_k$ defined with a trapezoidal membership function with parameters a, b, c and d , where a and d are respectively the minimum and maximum values from $(\Delta_\alpha)_k$, and b and c are respectively the values of the first and third quartile

of $(\Delta_\alpha)_k$. All the $(\delta_\alpha)_k$, $k = 1..K$ constitute a family of trapezoidal fuzzy sets, that can be integrated in the definition of the new learned fuzzy landscape regarding the direction α .

$$\hat{\mu}_{\alpha,d}^R(p) = t[\underbrace{(H_{\alpha_k} \circ \mu_\alpha^R)(p)}_{\text{direction}}, \underbrace{(\delta_{\alpha_k} \circ \mu_{\tau_R}^R)(p)}_{\text{distance}}] \quad \forall p \in S \text{ s.t. } \mu_\alpha^R \in [i_k, i_{k+1}[\quad (11)$$

where t is a t-norm fuzzy conjunction operator. A point p fits with the learned landscape $\hat{\mu}_{\alpha,d}^R$ if it fits with the directional information carried by $H_{\alpha_k} \circ \mu_\alpha^R$ and the related distance model carried by $\delta_{\alpha_k} \circ \mu_{\tau_R}^R$. Figure 7(a) gives an illustration of the resulting learned fuzzy landscape for the object R . This figure is to be compared with figure 3(b), where no distance model information was included. Integration of distance models as described for point of view *right* is repeated for each considered direction, and the four fuzzy landscapes can be intersected just as in equation (8). Figure 7(b) illustrates the intersection result, to be compared with the one obtained in figures 4(e) (without distance) and 6(b) (with global distance integration).

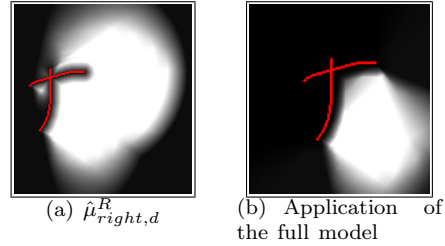


Figure 7 Learned fuzzy landscape with embedded distance for *right* point of view, and result of application of a full model with embedded distances (b).

6 Application to the recognition of handwritten gestures

















In order to evaluate the quality of positioning description provided by the models presented in this work, we set up an experiment aiming at recognizing handwritten objects.

6.1 Experimental datasets


Experiments were conducted on two distinct datasets of online handwritten gestures.

The first database (IME-OnDB) consists of a set of 5,525 handwritten records, collected from a handwritten input method on a pen-enabled PDA. 15 different writers participated in the collect. Each *record* is a pair of handwritten objects: a reference (it is one letter from the Latin alphabet), and an associated argument *gesture*. A gesture can be a punctuation mark, a diacritical, or a predefined

Table 2 Visualization of 16 selected classes from IME-OnDB database (a) and 16 classes from HCC-OnDB database (b).

| | | | |
|---|--|---|--|
|  acute |  grave |  circumflex |  dieresis |
|  accent deletion |  cedilla |  cedilla deletion |  case switch |
|  character deletion |  last character deletion |  caret return |  insertion |
|  space |  apostrophe |  comma |  dot |

(a) Classes of gestures included in IME-OnDB



(b) Classes of strokes spatial relations in HCC-OnDB

handwritten gesture for triggering an edition command in the input method (character deletion, case switch, space...). The argument is positioned with respect to the reference according to spatial relations that can be seen as examples of spatial relations classes, just like the *superscript* example cited above. The dataset is available from the IMADOC research group website^a, and is presented in more details in [Bou09]. Table 8(a) illustrates classes of gestures included in the database. Since examples from different gesture classes can have a similar shape (e.g. *comma* and *acute accent*), modeling their positioning with respect to the reference letter is necessary for a satisfactory recognition. The second dataset is an excerpt from an on-line handwritten Chinese characters database collected and provided by CASIA laboratory (Chinese Academy of Sciences, Institute of Automation). Original database contains thousands of classes of characters written by 60 different people. More details about this dataset are given in the works of Ma and Liu about Chinese character recognition [ML08]. For our experiment, we defined 17 classes of pairs of strokes extracted from simple characters. For each class, we isolated about 120 samples (2 per writer), where a sample is a pair of objects, one labeled as the reference and the other being the argument. Reference objects are made of one or two strokes, while the arguments are always a single stroke or a substroke. Table 8(b) illustrates examples of classes retained in our dataset (reference strokes are in red, argument strokes are in blue). In the sequel, we will refer to this dataset as HCC-OnDB.

6.2 Experiment design

Writer-independent cross-validation scheme

The experiment consists in classifying data samples using standard Support Vector Machine (SVM) classifiers, with different sets of positioning features combined with a fix set of shape features. All the experiments are writer-independent: training data samples (used for training spatial relation models and classifiers) and test data samples are from different writers. For the IME-OnDB database, 15 folds

^a<http://www.irisa.fr/imadoc/web/>

are defined (each fold corresponding to the data from one writer), and recognition rate for each set of features is averaged from a 15-fold cross validation. In the HCC-OnDB database, 15 folds are formed, each fold corresponding to the data from 4 writers, and the same writer-independent cross-validation process is applied to evaluate the average recognition rates.

Classifier parameters

The SVM classifiers have a Gaussian kernel, and their two parameters are optimized independently for each fold of each experiment by a grid search based on a random 10-fold cross validation over the training set. Cost parameter C ranges in $10^{-1}, \dots, 10^4$, and γ parameter ranges in $10^{-4}, \dots, 10^1$.

Feature sets

Different sets of spatial positioning features are utilized for recognizing objects. Two baseline sets are defined for comparing our method to commonly used positioning features. Feature set b is based on bounding box measures (such as horizontal and vertical distances between top, bottom, left and right boundaries of the bounding boxes of the two objects and distance between their center point). We compute histograms of angles between the objects and quantize angles into 18 bins, providing 18 features for baseline feature set c .

Features sets d and e evaluate the power of direct morphological directional and distance description. Mean adequacy degrees are evaluated in the four main directions by application of equation (3), constituting the four features of set d . A mean adequacy measure of distance is added in feature set e .

Finally, the three sets of features f , g , and h are composed of mean adequacy degrees computed for each of the N models learned by the method proposed in this paper, where N is the number of classes in each database ($N = 18$ for IME-OnDB, and 17 for HCC-OnDB). In other words, these feature sets result from the evaluation of positioning of the two objects in a pair sample calculated with respect to each learned model. Purely directional models are used for feature set f , global distance is incorporated for feature set g , and direction-wise distance modeling is integrated for feature set h . All the models are built with 8 bins histograms ($K = 8$, see paragraph 4.1). The t-norms used in models exploitation are the product t-norm for both landscapes intersection (see equation (8)) and combination of distance and direction in hybrid models (see equation (11)).

6.3 Results and discussion

Table 3 presents the different features sets and the recognition rates obtained on the two different datasets, average from 15-fold writer-independent cross-validation. Associate table 4 presents the statistical significance of pairwise feature sets comparison. We evaluated 5% level statistical significance by one-tailed paired student t-test over the 30 experiments conducted with each feature set (15 folds on each database). Justification for this statistical test is that the k-folds are writer-independent, and thus the standard deviation between folds is not meaningful. We consider the k-folds as different datasets (even if most of the training set is common between them). As shown in table 4, we test the hypothesis that each feature set

Table 3 Experiment results. Recognition rates are average on k-fold validation (k=14) for each database

| | <i>features</i> | % (IME) | % (HCC) |
|----------|---|---------|---------|
| <i>a</i> | (no positioning features) | 55.35 | 66.30 |
| <i>b</i> | 9 bounding box based features | 96.18 | 91.03 |
| <i>c</i> | 18 features angle histogram | 96.38 | 93.09 |
| <i>d</i> | 4 directional morphological features | 96.47 | 93.09 |
| <i>e</i> | 4 directional + 1 distance morphological features | 96.67 | 94.12 |
| <i>f</i> | N learned models (no distance) | 96.72 | 94.12 |
| <i>g</i> | N learned models (global distance) | 96.72 | 94.44 |
| <i>h</i> | N learned models (direction wise distance) | 96.79 | 94.85 |

Table 4 Statistical significance calculated with pairwise t-test at a 5% level, taking into account results from the two datasets. A checkmark in line x and column y indicates that feature set x is significantly more performant than y for recognition task.

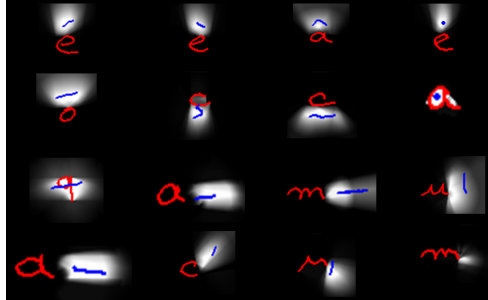
| | | | | | | | | |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | <i>a</i> | | | | | | | |
| <i>b</i> | ✓ | <i>b</i> | | | | | | |
| <i>c</i> | ✓ | ✓ | <i>c</i> | | | | | |
| <i>d</i> | ✓ | ✓ | - | <i>d</i> | | | | |
| <i>e</i> | ✓ | ✓ | ✓ | ✓ | <i>e</i> | | | |
| <i>f</i> | ✓ | ✓ | ✓ | ✓ | - | <i>f</i> | | |
| <i>g</i> | ✓ | ✓ | ✓ | ✓ | - | - | <i>g</i> | |
| <i>h</i> | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | - | <i>h</i> |

from b to h is actually better than less performant feature sets (according to the average recognition rates from table 3), and validate its significance if the 5% level is validated by the t-test. For instance, the fourth line of the table shows that d is significantly a better feature set than a and b , whereas no such conclusion can be drawn from its comparison with c .

Experiment a , where no positioning features are included in the description, confirms the need for a modeling of the relative positioning for recognizing classes in both datasets. Recognizing an editing gesture without description of its position only allows recognition of about half of the samples. In experiment b , a simple bounding boxes based spatial description is included, leading to a significantly increased recognition rate. Histogram of angles (feature set c) performs better than bounding box based features.

Experiments d and e confirm that the morphological description framework provides powerful features for spatial positioning (as we had already observed in [BMA06]), performing at least as good as histograms of angles, and significantly better than the two baseline approaches when a distance measure is included (feature set e).

Performance of models described in this paper appears in experiments f , g and h . First, feature set f performs significantly better than d , meaning that learning of models allows a better description of positioning compared to direct morphological

Table 5 Visualization of 16 learned spatial models corresponding to classes from IME-OnDB

description. It also performs better than the two baselines approaches. Likewise, learning models with embedded distance (h) significantly improves the recognition in comparison with direct morphological features with distance (e). However, no significant difference is noted between the global and direction wise distance integration (g and h). Actually, a closer look at the results reveals that distance integration only have a small impact on the IME-OnDB database. This is not surprising since the most confusing classes are the different types of accents (acute, grave, circumflex, and the accent deletion gesture) that differ not by their distance to the reference, but by their intrinsic shape. If we limit our statistical significance test to the HCC-OnDB database, it appears that embedded distance is significantly better than global distance integration at a 5% level. In general, the integration of embedded distance significantly outperforms purely directional models (comparison between h and f).

Facing with two distinct datasets confirms to some extent the genericity of our modeling approach. Expressivity in terms of spatial relations ranges from remote, disconnected objects (most of the classes from IME-OnDB) to more intricate relations such as *character deletion* class from IME-OnDB, several intersecting cases in HCC-OnDB, or a surrounding relation in the case of *case switch* class in IME-OnDB. Objects of different graphical complexity are properly handled by the models (substroke, single stroke or multi stroke objects). Table 5 presents visualization of spatial templates learned (with embedded distance) for 16 classes from IME-OnDB. Models adaptivity to varying reference shapes should be emphasized, especially for classes from IME-OnDB. Most classes from this database involve very different references shapes. For example, samples from *apostrophe* class admit c, d, j, l, m, n, s, t as reference letters. The *apostrophe* positioning model is then trained and exploited with very different reference letters, and the visual representation given in table 5 shows its good behavior.

7 Conclusion

In this paper we introduced a new general formalization for the modeling of spatial relations, with the novelty of addressing three major concerns of spatial positioning methods. First, thanks to their fuzzy morphological foundations, the models

consider the actual shapes of objects in describing their relative positioning. Secondly, the models are not restricted to *evaluation* of spatial relations, but can also handle *localization* task, in the sense that they can exhibit the region of the plane where the relation holds to a certain degree. Thirdly, the models can be automatically trained from data, allowing to capture the variability of spatial relationships between training objects. These three properties make the models eligible for dealing with handwritten objects for expression recognition and the experiments prove the superiority of spatial description in comparison to baseline approaches. Generalization capacity of our models was demonstrated by experimenting over different types of handwritten objects: editing gestures and accentuation marks on Latin characters, and strokes extracted from Chinese characters. We highlighted the benefit of distance integration into the models for the task of classifying handwritten objects. Future work directions include the exploitation of localization ability for assisting the segmentation task in structural processing of more complex handwritten patterns such as Chinese characters, or mathematical expressions.

Acknowledgments

This paper is a revised and expanded version of a paper entitled *Learning spatial relationships in hand-drawn patterns using fuzzy mathematical morphology* presented at the second international conference on Soft Computing for Pattern Recognition (SoCPaR), held in Cergy-Pontoise (France) on 7-10 of December, 2010.

The authors are grateful to Professor Cheng-Lin Liu from Chinese Academy of Sciences for providing them the HCC-OnDB database used in the experiments.

References

- [AMVG10] A. Awal, H. Mouchere, and C. Viard-Gaudin. Improving online handwritten mathematical expressions recognition with contextual modeling. In *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*, pages 427–432, nov. 2010.
- [Blo99] I. Bloch. Fuzzy relative position between objects in image processing: a morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):657–664, 1999.
- [Blo05] I. Bloch. Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, 23(2):89–110, 2005.
- [BMA06] F Bouteruche, S Mac, and E Anquetil. Fuzzy relative positioning for on-line handwritten stroke analysis. In *Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition*, pages 391–396, 2006.

- [BMK04] R. Bondugula, P. Matsakis, and J.M. Keller. Force histograms and neural networks for human-based spatial relationship generalization. In *Proceedings of Int. Conf. on Neural Networks and Computational Intelligence*, 2004.
- [Bou09] F. Bouteruche. *Description et exploitation du contexte spatial pour l'interprétation de tracs manuscrits : application l'interprétation de gestes graphiques dans les mthodes de saisie de texte*. PhD thesis, INSA de Rennes, 2009.
- [BR03] I. Bloch and A. Ralescu. Directional relative position between objects in image processing: a comparison between fuzzy approaches. *Pattern Recognition*, 36(7):1563–1582, 2003.
- [CA07] R.C. Cinbis and S. Aksoy. Relative position-based spatial relationships using mathematical morphology. In *Proceedings of the IEEE International Conference on Image Processing*, volume 2, pages 97–100. IEEE Computer Society, 2007.
- [Fre75] J. Freeman. The modelling of spatial relations. *Computer Graphics and Image Processing*, 4(2):156–171, 1975.
- [Gad97] P.D. Gader. Fuzzy spatial relations based on fuzzy morphology. In *Proc. Sixth IEEE International Conference on Fuzzy Systems*, volume 2, pages 1179–1183 vol.2, 1997.
- [JPPS10] R. Jeansoulin, O. Papini, H. Prade, and S. Schockaert, editors. *Methods for Handling Imperfect Spatial Information*, volume 256 of *Studies in Fuzziness and Soft Computing*. Springer Verlag, 2010.
- [KK10] D. H. Kim and J.H. Kim. Top-down search with bottom-up evidence for recognizing handwritten mathematical expressions. In *Proceedings of the 12th International Conference on Frontiers in Handwriting Recognition*, 2010.
- [MA04] S. Marukatat and T. Artières. Handling spatial information in on-line handwriting recognition. In *Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition*, pages 14–19, 2004.
- [Mat02] P. Matsakis. Understanding the spatial organization of image regions by means of force histograms: a guided tour. In P. Matsakis and L. M. Sztandera, editors, *Applying Soft Computing in Defining Spatial Relations*, volume 106 of *Studies in Fuzziness and Soft Computing*, pages 99–122. Physica-Verlag, 2002.
- [ML08] L.L. Ma, , and C.L. Liu. A new radical-based approach to online handwritten chinese character recognition. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08)*, 2008.
- [MS02] P. Matsakis and L.M. Sztandera, editors. *Applying soft computing in defining spatial relations*, volume 106 of *Studies in Fuzziness and Soft Computing*. Springer, 2002.

- [MW99] P. Matsakis and L. Wendling. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):634–643, 1999.
- [MWN10] Pascal Matsakis, Laurent Wendling, and JingBo Ni. A general approach to the fuzzy modeling of spatial relationships. In Robert Jeansoulin, Odile Papini, Henri Prade, and Steven Schockaert, editors, *Methods for Handling Imperfect Spatial Information*, volume 256 of *Studies in Fuzziness and Soft Computing*, pages 49–74. Springer Berlin / Heidelberg, 2010.
- [PS00] R. Plamondon and SN Srihari. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):63–84, 2000.
- [WNM06] X. Wang, J.B. Ni, and P. Matsakis. Fuzzy object localization based on directional (and distance) information. In *Fuzzy Systems, 2006 IEEE International Conference on*, pages 256–263. IEEE, 2006.
- [ZBC02] R. Zanibbi, D. Blostein, and J.R. Cordy. Recognizing mathematical expressions using tree transformation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1455–1467, 2002.
- [ZBZ05] L. Zhang, D. Blostein, and R. Zanibbi. Using fuzzy logic to analyze superscript and subscript relations in handwritten mathematical expressions. In *Proceedings of the 8th International Conference on Document Analysis and Recognition*, pages 972–976, 2005.